Data Security Optimization at Cloud Storage using Confidentiality-based Data Classification

Dorababu Sudarsa[1], Dr. A. Nagaraja Rao[2], Dr. A. P. Sivakumar[3]
Department of Computer Science and Engineering, JNT University Anantapur, Ananthapuramu Andhra Pradesh, India[1]
Department of Computational Intelligence-School of Computer Science & Engineering (SCOPE), VIT, Vellore[2]
Department of Computer Science and Engineering, JNTUA College of Engineering, Ananthapuramu, Andhra Pradesh, India[3]

***Abstract*—Data is the most assets for any organization, stored either in individual systems, server, or cloud platform. Cloud, one of the trending storage systems being adapted now a day is the state-of-the-art of the advanced technology. The major concern with this technological growth is privacy and security of data. Hoisting of data in this platform must be with privacy and security. Hence, there is an urge for service that provides security associated with data to the stake holders. Though the existing security for the data is provided at different levels incurred high cost in terms of processing time. This research aims at providing novel classification-based security algorithm (CBSA) composed with confidential-based classification and encryption with low cost. The confidential-based classification classifies the data into three different levels based on its degree of confidentiality; confidential-based encryption applies a suitable and proportional security mechanism dynamically to each of the levels of data. Thus, the data security process will become optimal and cost effective. The proposed algorithm has outperformed the existing algorithms in terms of processing time and entropy. The processing time and entropy of proposed algorithm has improved by 10%.**

Keywords*—Cloud storage; data privacy; data security; data classification; degree of confidentiality

I. Introduction

Data is a most important asset of organizations which may in unstructured or structured form. Irrespective of type of data preserving the data sensitivity is very important for an organization to provide the promotional security to it.

With the advancement in technology today cloud has been a buzz word with it versatile services such as storage, sharing of data, and security for many applications. This is because of its flexibility, reliability; economic scalability, Cloud Service Provider (CSP) interaction [1], cost effectiveness and maintenance free [2]. Since the client data is managed by third party vendors security has been a concern over the platform. Cryptographic system [3] is common security adapted via cloud these days. Cloud Storage is an ideal model of online networked storage comprises multiple virtual servers [4] which facilitate the users to store the data remotely, to access them at any time and from anywhere as shown in Fig. 1. The main benefit of using the cloud storage is that it is possible to store the data as required quantity depends on the business needs. But to minimize the storage, manage and processing time (PT) and also to minimize the energy usage [5], one of the solution may be reducing the duplication of the data but also simplifying the storage process and adapting the flexible and suitable framework.

According to the preliminary study data that is accessible to the general public can be as low-level secure data. And certain data or information pertaining to the military, financial, intelligence agency, or police secret operations is regarded as extremely confidential and can consider as high-level secure information since it requires high-level. In this scenario the level of security algorithm has major role in terms of significant considerations like speed, effectiveness, efficiency and cost. Low security measures may put sensitive data at risk and expose us to cybercriminals if we adopt them for cost-saving reasons. If we apply high security mechanism, it works effectively and very useful for vital data, but may not be cost effective for the no security challenging data. Frank Simorjay et al. defined some parameters or factors in [6] that can be used to achieve data security and confidentiality, including data access control, authorization, authentication, etc.

The easiest and most practical security measure is the classification-based system. The level of security for data varies depending on the needs of the data owner and the type of data; security algorithms are also used in applications and are measured for speed, efficiency, cost, and energy consumption. In this respect, it is not ideal and accurate to encrypt an application's data using a single encryption method. There so many benefits [7] providing by the cloud based on the data classification process, and can be achieved key issues such as data privacy, integrity and accessibility and also at any cost it should be easy to use and less expensive [8].

Hence, data security is also important while transmitting it between the cloud servers [9] and various solutions are designed for it. Even to secure the data at various levels of cloud many methods were proposed on diverse domains such as storage, access control, network, software, hardware, hypervisors; classification mechanism or technique to classify the data just before it store into any related encryption system is mandatory. Since, data storing into the cloud is not has equal sensitivity level or degree of confidentiality; if entire data is encrypted by using single algorithm may lead to deficiency of security. Even present classification methods used to achieve security; are not use the high security mechanisms to provide higher degree of security. Therefore, it is compulsory and essential to optimize the data security in terms of security mechanism used, computation cost, and the resources consumption. So, in this paper a novel algorithm called CBS (Classification based security) is proposed that

consist of sub-algorithms, find DC to find the degree of confidentiality of every attribute, CBC(confidential-based Classification) algorithm to classify the data based on its confidentiality level, CBE(confidential-based encryption) algorithm to encrypt the data by using light weight security mechanisms, CBD (confidential-based decryption) algorithm to decrypt the data by using the same light weight security mechanism which is encrypted.

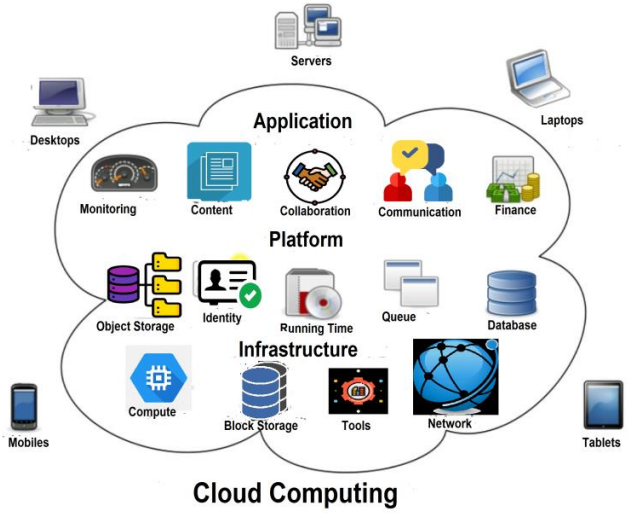

Fig. 1. Cloud computing environment.

A. Contribution of the Work

The contribution of the proposed work is providing the security to the data before it stores onto the cloud. As well as the security process should be less complex. It is achieved using the algorithm named as CBS (confidentiality based Security) that comprises the sub-algorithms called findDC, CBC, CBE and CBD. The proposed algorithm performs the following actions to protect the data:

1) Consider the data set D that to be protect, that has the attributes $A_1, A_2, \ldots\ldots A_{n-1}, A_n$

2) Find the Degree of Confidentiality (DC) of each attribute of D.

3) Classification of data by using Confidentiality-based classification (CBC) algorithm in terms of their attribute's DC value.

4) Now, apply the suitable security mechanism dynamically to each class of data part according their confidentiality level. Here RSA algorithm is used for the basic or low confidential level data to secure, CP-ABE algorithm for the medium confidential level data and CP-ABE-SD algorithm for the high confidential data to protect the data.

5) Transport layer security (TLS) protocol is used to transmit these three class types of data to store onto the cloud since TLS is also a good secured transmission protocol.

6) Decrypt the data which is requested by the user to read, work or update from the cloud by using the same security mechanism used in encryption based on its confidentiality level.

B. Organization of the Sections

Left over part of this paper is organized as follows: different solutions for the given problem and their drawbacks are specified in Section II. The selection of better security algorithm is needed for the proposed algorithm that is given in Section III. The proposed work is clearly described with the system model in Section IV. The results and its discussion is shown in the Section V and Section VI respectively. Finally, Section VII concludes the paper.

II. Related Work

The recent work done on the data security to provide using various mechanisms are given here through the thorough survey. Sandeep K. Sood [10], focused on the security of data both at cloud storage and in transit mode by proposing a framework with two phases, one deals about the transfer process and secure data storage in the cloud and second one deals about the data repossession process from the cloud and creation of requests for the data access, accreditation of the digital integrity and signature, double authentication. They performed data classification according to the CIA triad parameters. But, calculating the SR (sensitive rating) from the values of C, I and A is a too delay process, instead we can take SR value directly either from customer or by using weight allocation techniques.

In study [11] Munwar Ali Zardari et al, proposed a confidential based data classification model that classifies the data as classes only that is sensitive and non-sensitive data by using machine learning technique called K-NN classifier to ensure data confidentiality. Among these, sensitive data need to be more security hence RSA algorithm is used and non-sensitive data is stored directly in the cloud servers. However, this type of classification of data is not given optimal solution.

In study [12] MingLi et al. proposed a patient-centric framework to control access of Patient Health Records (PHR) which are deposited in semi-trustable servers. Attribute-based Encryption (ABE) technique is used to protect the PHRs, hence achieved fine-grain and scalable data access control. Still PT for encryption and decryption is taking more by this model. Yuan Cheng et al. [13] proposed a framework to restrict data access by the third party applications (TPAs). Applied some policies for restricting the data access and hence, the data privacy is attained from the TPAs. This framework can give only data confidentiality, but not ensures the data security.

Data classification making at various phases in the social networks is presented by Sergio Donizetti Zorzo et al. in study [14] that classified the data with respect to the security parameter called confidentiality in the network. In study [15] Dr. N. Srinivasu et al. identified the threats at different levels in the cloud and addressed the security requirements to resolve them by using the security mechanisms. The requirements include data encryption, confidentiality, data integrity, data authorization and authentication and data privacy. And mapped those requirements to different cloud services to obtain the coherence and integrity. But they did not specified solution for those issues they outlined.

In [16] Sudarsa, D. et al. proposed a method that identifies diverse kind of data, attributes, their sensitivity level and then classified the data into small parts to store them into the cloud as number of clusters. Thus, data accessing is very easy by using suitable access rights. The security levels are defined in the cloud as per the data content type and accessibility. But with this data will not classified systematically and not feasible for huge data.

In study [17], Rasmeet Kour et al. presented a data classification technique to classify the data into sensitive i.e. and non- sensitive. Then sensitive data is protected by using Blowfish algorithm and whereas non-sensitive data is stored straight away into the cloud without encryption; hence processing overhead and time is reduced. The secure cloud system upgraded, by divided the cloud into segments, data also divided and then stored them into those segments, but not entire data onto a single cloud. But, they are not considered all the criteria which provides better security, for instance, some data may not be sensitive or non-sensitive. In that case still better classification process is required.

In study [18], data is classified in respect of the three factors of CIA Triad: Confidentiality, Integrity and Availability. It takes the data as input and produces the data as three categories such as public, private or restricted data as per the value of Sensitivity Rating, the above three parameters collectively termed as Sensitivity Rating. But C, I and A values should be taken from the customer or by using any weight allocation mechanisms; it leads to extra burden to the system. In [6], Frank Simorjayet al. given the data classification method based on the three properties: Content, Storage and access control. Every property divided again into sub properties.

In study [19] authors proposed a model that classify the data using the fuzzy logic. It characterizes the data based on the security requirements of the owner using CIA triad of information security system and then user data is classified by using the fuzzy logic theory based on CID triad. However, designing and implementing the fuzzy logic systems is also somewhat difficult. In [20], Rizwana Shaikh et al, identified some set of factors which supports data classification in the cloud, analysed tem and security levels are defined with respect to the type of data and its accessibility as per the required confidentiality and access restrictions. But, they are considered the parameters in different angles which may not give the robust solution for the data security issue.

Ahamad, Danish, et al. in study [21], A privacy preserving model is implemented for cloud division that involved with two steps such as data sanitization and restoration by utilizing an optimal key generation. Here key optimization is done using J-SSO algorithm by developing a multi-objective function comprised with three factors: information preservation ratio, hiding ratio and degree of modification. Though a good algorithm is proposed, the security enhancement is needed more. Tawalbeh, Lo'ai, et al. in study [22], an efficient framework named confidentiality-based cloud storage is proposed that can assures integrity and confidentiality through data classification and reduces the PT by applying TLS, AES and SHA based on type of data classified. But they did not considered data classification based on the customer needs or professional body's suggestions to enhance and not used asymmetric public key like RSA and ECC which can provide greater degree of security.

DVK Vengala et al. [23] proposed an authentication algorithm with three factors and a secured ECC based data transfer method to transfer the data and to check the authentication of the user for accessing the data securely. Though, they given solution for secure data transfer with distributed cloud servers, security enhancement required for the present scenarios. In study [24], proposed an architecture that provide the security for our data while any unauthorised data is trying to access by using basic simple algorithms, but those are not sufficient to face the present security scenarios.

In study [25], the authors provided a secure cloud storage construction called FABECS using fully ABE approach, and CP-ABSE and DET-ABE constructs are as main building blocks of it. But this approach is provided security with little bit complexity, and not supporting the group attribute data securing. In study [26] M. Thangavel et al. proposed an integrity verification framework for cloud storage security based on Ternary Hash Tree (THT) and Replica based Ternary Hash Tree (R-THT), which will be used by TPA to perform data auditing. It performs Replica-level, File-level and Block-level auditing with tree block ordering, storage block ordering for verifying the data integrity and ensuring data availability in the cloud. The framework supported error localization with data correctness, dynamic updates with block update, insert and delete operation also. The structure of THT and R-THT tried to reduce the computation cost and to improve the efficiency in data updates. However still need to reduce the computation cost.

In study [27], Fursan Thabit et al. proposed a two layer encryption to improve cloud computing security, one works based on Shannon's theory of diffusion and confusion with logical operations such as XOR, XNOR, and shifting by dividing the plaintext and key into equal parts, and another one works based on structures of genetics based on the Central Dogma of Molecular Biology for cryptographic. But it suffered with space complexity and still needs robust security on the cloud storage.

In study [28], O. Arki et al. presented a CID triad model with fuzzy logic to classify data to provide the security according to requirement of the security. But not achieved the data security up to the level for the present scenarios. In [29] A. Yeboah-Ofori et al. presented an encryption mechanism to discover data security by merging AES algorithm, cloud storage, and Ethereum smart contracts in the cloud AWS S3.to improve the blockchain security in the cloud. But not satisfied all security parameters with that mechanism.

In study [30] P. Swathika et al. developed a method to enhance cloud storage security by interlinking advanced client-side encryption with (RBAC) Role-Based Access Control by dynamically grouping the users into predefined roles with related permissions. Even though, they are unable achieve the data security up to the mark. In study [31], R. R. Prasad et al. proposed a method named, Balanced Genetic

Algorithm (BGA) to enhance the data security, scalability, and decouple the data life cycle from the core encryption process. But this method is sufficient to face the present scenarios.

In study [32], Mahesh Muthulakshmi R et al. proposed system named the Weight-Improved Particle Swarm Optimization Algorithm (WI-PSO) and machine learning classifiers to enrich the data security in the cloud. Still system need to enhance the security with light weight algorithms with less PT. In [33], Suchitra R et al. proposed a technique called fragment-based encryption that utilizes an algorithm to generate variable length different keys based on the required confidentiality level for the document fragment. Still the mechanism not provided high security for the cloud data.

In study [34], N. Dwivedi et al. used fully holomorphic encryption in their work that enables computations on encrypted data without having to first decrypt it. This makes it possible to process the sensitive data securely in the cloud, to preserving privacy and confidentiality of the data. Still processing cost increased little bit with this mechanism.

By observing the above survey, it is understand that there is urge to develop an optimized framework and the suitable algorithm to improve the security and efficiency by reducing the PT of the overall process of the system. In this paper, we focused on optimization of data security by considering the good framework and suitable algorithms which supports that framework, that too light weight and advanced security algorithms to improve the secure cloud storage by reducing the PT of their encryption and decryption tasks. Our work offers the following unique and better security features for their data while owner or user wanted to store and access their data:

1) Computing the classification parameter.

2) Classification of data into three classes by using the Degree of Confidentiality value.

3) Selection of the better and light weight security algorithms useful in the proposed algorithm by comparing different existing algorithms.

4) Each class of data is securing with the proportional level security algorithm as given below:

a) Low confidential data secured by using low level security algorithm.

b) Moderate-confidential level data secured by using moderate level security algorithm.

c) High confidential data secured by using high level security algorithm.

5) Performance Evaluation of proposed algorithm by comparing with existing algorithms in terms of their PT and average entropy.

III. Selection of Light Weight and Optimal Security Algorithms for the Proposed Framework

To select and use light weight optimal security algorithms in the proposed framework, different security algorithms are compared in terms of various parameters. For this, we have executed and compared the performance of different security algorithms like RSA[35], IBE[40], ABE[41], KP-ABE[42], CP-ABE[36], Enhanced CP-ABE[43], FABECS[25], and CP-ABE-SD[37] algorithms in respect of their PT of encryption and decryption process, entropy per byte of encryption and in terms of memory used. All these algorithms are implemented in java Eclipse IDE and used supporting packages such as java security and java crypto. These packages provide security features such as key management infrastructure, key generation, authentication and authorization, encryption, decryption. Each algorithm developed in java, transformed into a jar file and then included that jar to crypto library externally. Text files of sizes 25KB, 50KB, 1MB, 2MB, 3MB, 4MB and 5MB are used as input to the encryption process. Each output file in the encrypted form is saved, that is taken as input for decryption process. To analyse thru the comparison, same files are used for all the algorithms as input during the course of the experiment. All these implementations and analysis works are carried out in the same system; hence processor and memory conditions remain standing as it is for all the algorithms for the comparison.

By the observation of above algorithms, it is understood that above said all algorithms are asymmetric and RSA algorithm takes less time for encryption than IBE, ABE, KP-ABE, CP-ABE, Enhanced CP-ABE, FABECS, and CP-ABE-SD. And CP-ABE, Enhanced CP-ABE, FABECS, and CP-ABE-SD are advanced and high-security algorithms. Here, light weight and efficient algorithms are selected for the obtained three levels of data to secure. Hence, RSA algorithm is used in our proposed algorithm for the low-confidential level data to secure and CP-ABE algorithm for the medium-confidential level and CP-ABE-SD algorithm for high-confidential level data as a part of the work to achieve the minimal PT.

IV. Proposed System

The main aim of this work is to resolve two concerns that user happenstance when utilizing the CC services. One is about the threats or hacking of data either externally or internally. Another one is without considering the degree of confidentiality, encrypting the entire data may be infeasible. For instance, suppose a 10MB data entire block encrypted using same key size and same security level mechanism, it may not be feasible. Because, it takes high PT to encrypt the entire data block.if we consider the degree of confidentiality, it is possible to save time by classifying the data based on its confidentiality level. After clssification process, if 10MB data is classified as 3MB as basic level data, 4MB as confidential data and 3MB as high Confidential data; it is better to apply the basic level security algorithm to the low confidential level data, moderate security mechanism to the confidential level data and high security mechanism to the high confidential level data. With this, the encryption and decryption PT of basic and confidential level data can be reduced and hence the overall PT for the entire data will reduce. Therefore, a well-defined framework and a novel security algorithm that supports the specified framework are needed. In this paper, we proposed such a framework and implemented a novel algorithm called CBS (confidentiality-based security) algorithm which supports it. CBS incorporates a classification algorithm and algorithms for encryption and decryption of the data based on its confidentiality level by using DC value of its

attribute. Here DC value of an attribute can be used to classify the entire data into three classes such as low confidential data, confidential data and high confidential data. Here, data classification is a process that permits individuals and organizations to classify all different types of data assets into different categories based on its confidentiality degree that determines the enhancement of data security needed. Classification guarantees the information sensitivity and hence proportional protection can be provided to each sensitive level of information.

To reduce the PT of encryption and decryption of the data, a novel data classification algorithm named as CBC is incorporated in the proposed algorithm CBS. The CBC classifies the data in three ways based on their attribute's degree of confidentiality such as Low Confidential, Confidential and High Confidential level. Low level data is low-sensitive, hence we can apply light weight security algorithm. As per the study done, the algorithms takes less memory and less PT in encryption and decryption, we applied RSA [35] algorithm at low-confidential level data, the CP-ABE [36] algorithm at medium-confidential level, and CP-ABE-SD [37] algorithm for High confidential data since it need high security. For transmitting the data we use TLS [38] algorithm. After storing the data in to the cloud securely, if any users want to access it, then decrypted using the same algorithm it was encrypted and then declassifies it to achieve this process the proposed system model is designed as given below Fig. 4.

A. Proposed CBS Framework

The Proposed system works in three main steps, one is Data classification, second one is Data Storage and third one is Data retrieving.

1) Data classification: To perform the data classification, first data is send to the machine learning algorithm to train as shown in Fig. 2. Once the data is trained, then data classification can be performed using same machine learning algorithm based on the trained data as shown in Fig. 3. So the output obtained is data in three different classes, Low confidential, confidential, and high confidential level data. After the classification, data is encrypted using corresponding security algorithm.

2) Data storage: The data is stored into the cloud server after the data encryption with corresponding security algorithm.

3) Data retrieving: When the user/owner anted data, then data can be retrieved from the cloud after data decryption and declassification. Now the data is obtained to the user in the original form.

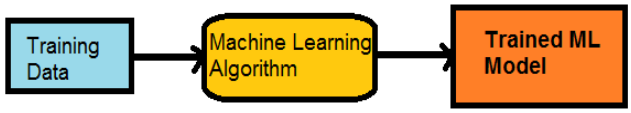

Fig. 2. Making trained machine learning model for data classification.

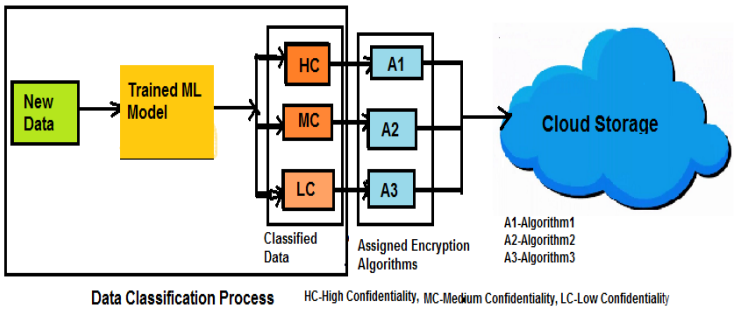

Fig. 3. Data classification process before storing on the cloud storage.

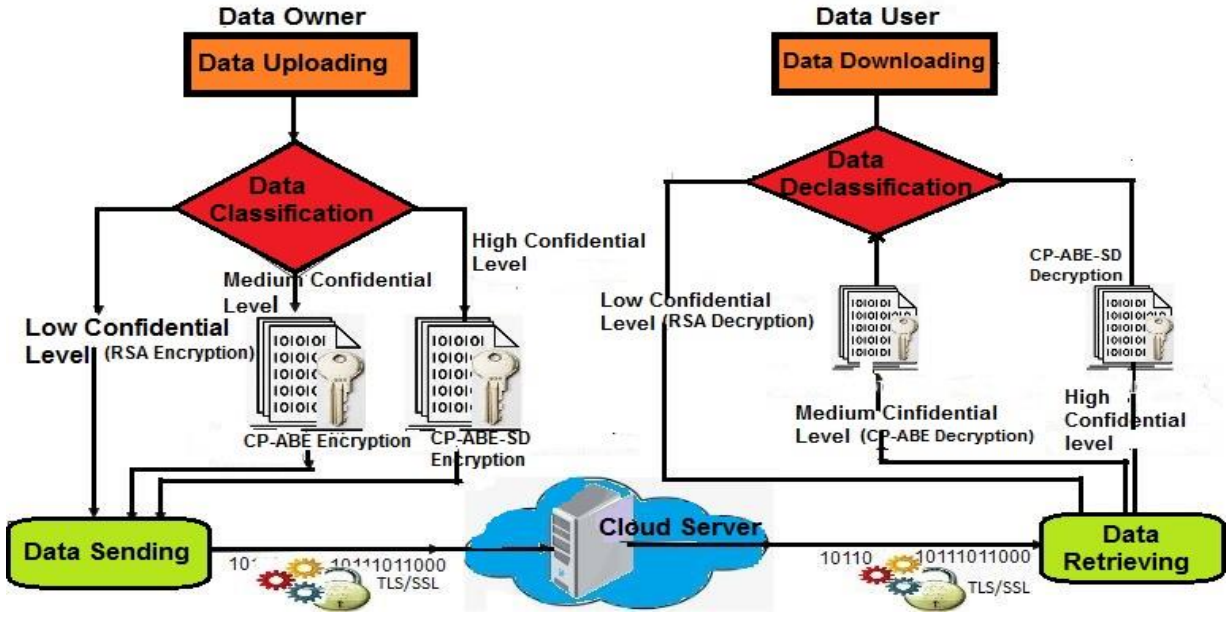

Fig. 4. Proposed CBS framework with three confidential levels.

The proposed model is shown in the Fig. 4 that shows data in three different security levels: Low Confidential, Medium-Confidential and High Confidential.

1) Low confidential level or basic level: The data which has very low level degree of confidentiality comes under the low confidential or basic level security. Low level security can be provided to this level or class of data like videos, photos and public data which is basically not requires high degree of security. Therefore, this level proposes only basic level of security and is used in online by the most of the organizations. Basic level data can be considered as low impact data. The loss of data integrity, availability or confidentiality of data in the cloud causes to have less or no contrary effect on the stake holders of the data. In the sense, even though the integrity, confidentiality or availability is compromised, less impact on its regular progression, financial loss or order of tasks. To provide low level security, RSA algorithm is used to encrypt and additionally TLS used to transmit data between the client's and server's applications by using HTTPS.

2) Medium confidential level: The data with moderate degree of confidentiality is comes under this confidential or sensitive level. The data like personal files, photos and videos, family data, friend's data are known as confidential data. Confidential level data can be considered as moderate impact data. The defeat of integrity, availability or confidentiality of data in the system or cloud, causes medium contrary effect on the stake holders of the data. That is, if the integrity, confidentiality or availability is negotiated, its effect is at moderate level on the flow of progression, order of tasks etc. But, financial loss should not be considered as moderate impact. Here, data is encrypted using CP-ABE algorithm at client side, and then transmitted thru the network.

3) High confidential level: In this level, data can be considered as very high impact data. This level will handle the most significant data such as financial transactions, criminal information, military information, business policies and patient health records. This type of data can be considered as high impact data. Stakeholders are very bothered about losing the data like this since it is high confidential, using all the new offered services are still avoided. Therefore, at this level the data will be with high degree of confidentiality, the user is provided high security to maintain high confidentiality and integrity by using CP-ABE-SD algorithm that guarantees the integrity and confidentiality of data. Data is encrypted using this algorithm before sending to the cloud servers so, user has assured that data was not damaged or tampered.

B. Proposed CBS Algorithm

To obtain the optimal solution for the data security at the cloud storage, a novel algorithm named as CBS (Confidentiality Based Security) algorithm is proposed. This work focused on structured data because, now a day's most of the data of organization will be in structured manner. So, this algorithm is completely works on the structured data that is the data which represents in two dimensional in rows and columns. The CBS algorithm works with the following steps:

1) The first and most important step is obtaining the DC for every attribute of the data set D by using FindDC(A_i) algorithm.

2) The data which outsourcing to the cloud is classified as Low-confidential, Medium-confidential, and High - confidential data, by using CBC algorithm.

3) Protection of data can be done for each class of data independently by using CBE algorithm that utilizes already existed, light weight and well defined encryption mechanisms. Here, RSA algorithm is used for low confidential level data, CP-ABE algorithm is used for medium-confidential data and CP-ABE-SD algorithm is used to protect high confidential data.

4) Now data is store onto the cloud, we can call now it as a 'Secure cloud'.

5) Whenever any user wants to access the data, that can be retrieved from all these three class types of data after performing the decryption process by using CBD algorithm that applies decryption mechanisms to the data encrypted.

The structured data can be imagining as a database in which data will be in set of rows and columns. To classify the data represented in structured manner, its attributes plays vital role. So, in this work, patient health records are taken as data set and classified the data as three levels as specified above based on their attribute's degree of confidential values.

Suppose, the data set D is represented as:

$D = [A_1, A_2, A_3, \ldots\ldots\ldots\ldots, A_{n-1}, A_n]$

where, $A_1, A_2, A_3, \ldots\ldots\ldots\ldots, A_{n-1}, A_n$ are attributes of the dataset or database

n is number of attributes in the database.

$W_i [A_1]$: DC value given to attribute A_1 by the professional members 1,2,3…….m

$W_i [A_2]$: DC value given to attribute A_2 by the professional members 1,2,3…..m

-

-

$W_i [A_n]$: DC value given to attribute A_n by the professional members 1,2,3…….m

From the value of $W_i [A_j]$, DC value for all attributes is calculated and represented as:

$DC[A_1]$: Degree of Confidentiality(DC) of attribute A_1

$DC[A_2]$: Degree of Confidentiality(DC) of attribute A_2

-

-

$DC[A_n]$: Degree of Confidentiality(DC) of attribute A_n

Now, based on the degree of confidentiality (DC) value obtained for each attribute A_i, the data is classified into three classes such as Low Confidential, Confidential and High

Confidential level data. The following Table I shows the DC value of an attribute and its confidentiality level:

TABLE I. DC VALUES OF DIFFERENT CONFIDENTIAL LEVEL DATA

Confidentiality Level	Degree of Confidentiality (DC) Value
Low-Confidential level data	0
Medium-Confidential level data	1
High-Confidential level data	2

The degree of confidentiality value is decided based on the values given by the different professionals and domain experts. In this work, a survey is conducted on different attributes of patient health records and collected the degree of confidentiality value for every attribute from more than seven fifty number of domain experts and different professionals by using the link https://tinyurl.com/pu3d3r4x. Then DC values obtained for all different attributes are used as trained data to classify the remaining data.

Algorithm to find the Degree of Confidentiality for attributes:

Input: $W_i[A_j]$: Value given to attribute A_j by the professional members i =1,2,3……..m
Output: $DC[A_j]$: Degree of Confidentiality of attribute A_j, where j= 1,2,3……..n

Algorithm: FindDC(A)
1. For j = 1 to n
   1.1 For i = 1 to m
   $Sum[A_j] = Sum[A_j] + W_i[A_j]$
   1.2 $DC[A_j] = floor(Sum[A_j] /m)$

The average value obtain for a particular attribute is assigned as its degree of confidentiality value. If the degree of confidentiality of an attribute A_i is 0 (zero), then that attribute is comes under Basic level or Low confidential level data. If the degree of confidentiality of an attribute A_i is 1 (one), then that attribute is comes under Confidential level data. Similarly, if the degree of confidentiality of an attribute A_i is 2 (two), then that attribute is comes under the High confidential level data. The data which is under a particular attribute will be comes under the same attribute's level. All the attributes which are having the same degree of confidentiality value will come under a class C_k. Where k=0, 1 or 2 represents the class type. So, in this context three classifications are formed such as C_0, C_1 and C_2 and they named as Low-confidential level data, Medium-confidential data and High-confidential data respectively. So, the attributes and its data in D will distribute to C_0, C_1 and C_2, This classification process is done using a multi-class classification algorithm called Decision Tree Algorithm [39].

For an instance,

D_i	A_1	A_2	A_3	A_4	- -	A_{n-2}	A_{n-1}	A_n
DC[Ai]	0	0	1	1	- - -	2	2	1

Now according to Degree of Confidentiality, the classifications are formed and expressed as:

C_0 : $[A_1, A_2, A_6]$

C_1: $[A_3, A_4, A_5, A_n]$

C_2: $[A_{n-2}, A_{n-1}]$

All the data of low confidential level will be in class C_0, all the data of medium confidential level will be in class C_1 and all the data of high confidential level will be in class C_2. This process is done using the following CBC algorithm.

Confidentiality based Classification(CBC) algorithm:

Input: Attributes set A_i in the data set D
Output: C_0, C_1, C2

CBC algorithm(D):
Begin
1. For i=1 to n
   1.1 If $DC[A_i] = 0$ then
   Add A_i to C_0
   1.2 Else If $DC[A_i] = 1$ then
   Add A_i to C_1
   Else
   Add A_i to C_2
End.

Once entire the data is classified into three different classes, and then applied the corresponding security mechanisms. This encryption process can be carried out using the CBE algorithm that applies the corresponding encryption algorithm to a particular class of data. The data in the class C_0 will be encrypted using Enc_RSA( ) since it has low confidential data, the data in the class C_1 is encrypted using the Enc_CP-ABE() algorithm since it has moderate confidential data and the data in the class C_2 is encrypted using Enc_CP-ABE-SD() algorithm since it has high confidential data.

Confidentiality Based Encryption (CBE) algorithm:

Input:C_0, C_1, C2
Output: $Cipher_1$, $Cipher_2$,

CBE (C_i) Algorithm:
Begin
1. For i=0 to 2
   1.1 If i = 0 then
   $Cipher_0$ = Call Enc_RSA(C_0);
   1.2 Else If i = 1 then
   Cipher1 = Call Enc_CP-ABE(C_1);
   Else
   $Cipher_2$ = Call Enc_CP-ABE-SD(C_2);
End.

Once data is encrypted, that can be retrieved only after its decryption is successful when the user is requested, otherwise data cannot be retrieved. The decryption process can be carried out using the CBD algorithm that applies the corresponding decryption mechanism to a particular class of data. A particular class C_i of data can be decrypted using the same algorithm which is encrypted with. That is, the data in the class C0 is need to decrypt using Dec_RSA() since it has low confidential data, the data in the class C1 is decrypted using the Dec_CP-ABE() algorithm since it has moderate confidential data and the data in the class C2 is decrypted using Dec_CP-ABE-SD() algorithm since it has high confidential data.

CBD (Confidentiality Based Decryption) algorithm:

Input: $Cipher_1$, $Cipher_2$,
Output: C_0, C_1, C2

CBD ($Cipher_i$) Algorithm:
Begin
1. For I = 0 to 2
   1.1 If i = 0 then
       C_0 = Call Dec_ RSA($Cipher_0$);;
   1.2 Else If i = 1 then
       C_1 = Call Dec_CP-ABE($Cipher_1$);
       Else
       C_2 = Call Dec_CP-ABE-SD($Cipher_2$);
End

The data can be sending to the network media directly because it is securing transmitting data by using the secured transmission protocol like TLS [30]. TLS is a security protocol that provides privacy, security and data integrity for the intercommunications via Internet. A main aim of TLS is to encrypt the intercommunication between servers and web applications like web browsers and also to encrypt other internet conversations like voice over IP (VoIP), messaging and email.

V. Experimental Results

The performance of the proposed algorithms are calculated in terms of PT of encryption and decryption process performed on the entire data by using only CP-ABE-SD algorithm, only FABECS algorithm and using our proposed CBS algorithm. A simulator is built to evaluate the proposed framework, which was developed using java Eclipse environment and used java security and java crypto packages; these packages features such as authentication and authorization, encryption, decryption, key management infrastructure and key generation. They also comprise the classes and interfaces needed to execute the java security architecture. CP-ABE-SD algorithm is implemented in java, converted into a jar and then included that CP-ABE-SD jar to crypto library externally. We have finished execution with all necessary verifications and validations. The simulation experiments is conducted under the same platform: Intel(R) Core(TM) i3-5005U CPU, 2.00GHz processor, RAM of 8 GB and Microsoft Windows 8.1 Pro.is the operating system used.

The encrypted each file is saved as a file and sent as input to the decryption process. For comparison, the same input files have been used for all algorithms throughout the experiment. All these implementations and analysis work are carried out in the same system; hence processor and memory conditions maintained same for all the algorithms. The experiment is done on data blocks with various sizes. Fig. 5, 6 and 7 show the performance evaluation of existing algorithms FABECS, CP-ABE-SD and proposed CBSA algorithm, when encrypting, decrypting the data blocks ranging from 10 MB to 100 MB and also their Average Entropy. The sizes of data blocks represented in the x-axis in megabytes and the PT in y-axis in seconds.

The Fig. 4 given below shows the performance evaluation of our proposed CBS algorithm with existing security mechanisms in terms of their encryption PT.

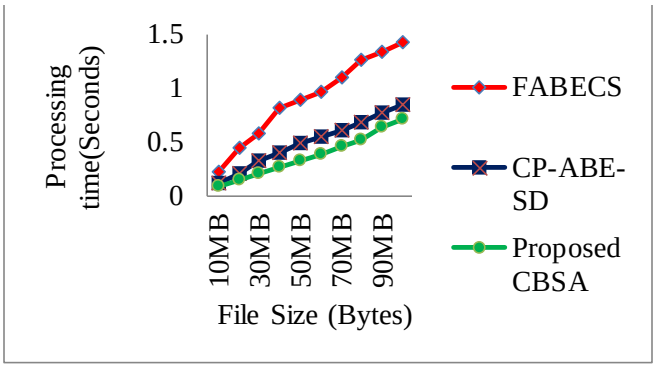

Fig. 5. Performance analysis of existing FANECS, CP-ABE-SD and proposed CBSA in terms of encryption PT.

The performance evaluation of our proposed algorithm is done in terms of PT of encryption by comparing the proposed algorithm with existing algorithms FABECS and CP-ABE-SD applied on the entire data. A 10MB data is encrypted by our proposed algorithm in 0.08 seconds, it is 63.63% is better than FABECS and 33% better than the CP-ABE-SD algorithm. A 50MB data is encrypted in 0.32 seconds; it is 64% better than FABECS and 33.33% better than CP-ABE-SD. In this case we can benefit 35.9% and 66.66% PT when compare to FABECS and CP-ABE-SD. Similarly a 100MB data is encrypted in 0.71 seconds; it is 50% better than FABECS and 15.47% better than CP-ABE-SD algorithm. In this case, we can benefit 50% and 86.52% PT when compare to the FABECS and CP-ABE-SD algorithms. The complete and clear performance evaluation analysis on the encryption PT given by FABECS, CP-ABE-SD and our proposed algorithm is shown in the Fig. 5 above for the different size of data files from 10MB to 100MB.

The Fig. 6 shows the comparison between performance evaluation of our proposed algorithm and remaining two security mechanisms in terms of their decryption PT.

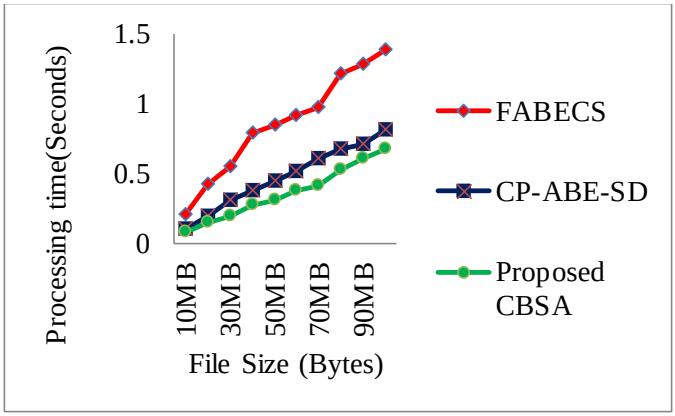

Fig. 6. Performance analysis of existing FANECS, CP-ABE-SD and proposed CBSA in terms of decryption PT.

The performance evaluation of our proposed algorithm is also done in terms of its PT of decryption process by comparing it with other alternative algorithms FABECS and CP-ABE-SD. For decryption also different size of data files from 10MB to 100MB are taken and applied decryption process of the same algorithm by which it is encrypted. A

10MB data is decrypted by our proposed algorithm in 0.08 seconds; it is 61.91% is better than FABECS and 20% better than the CP-ABE-SD algorithm. In this case we can benefit 38.09% and 80% of PT when compare to FABECS and CP-ABE-SD. A 50MB data is decrypted in 0.38 seconds; it is 63.53% better than FABECS and 31.12% better than CP-ABE-SD algorithm. In this case, we can benefit 36.47% and 68.88% PT when compare to FABECS and CP-ABE-SD. Similarly a 100MB data is encrypted in 0.67 seconds; it is 51.45% better than FABECS and 17.29% better than CP-ABE-SD. In this case, we can benefit 48.55% and 82.71% PT when compare to the FABECS and CP-ABE-SD algorithms respectively. The complete and clear performance evaluation analysis on the decryption PT given by FABECS, CP-ABE-SD and our proposed algorithm is shown in the Fig. 6.

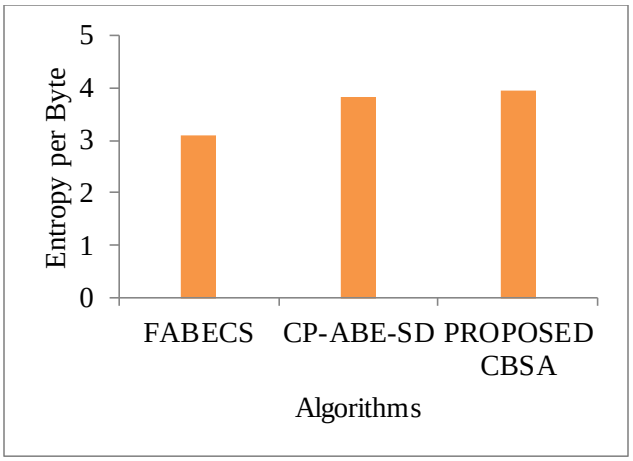

Fig. 7. Average Entropy per byte given by FABECS, CP-ABE-SD and Proposed CBSA.

The Fig. 7 shows that our proposed algorithm scores greater average entropy per byte of encryption. Entropy is the measure of degree of randomness of the information. if we apply only CP-ABE-SD algorithm, it gives average entropy per byte of encryption is 3.9349. if we use only FABECS for the entire data, it gives 3. 0958. But if we apply our proposed frame work and algorithm it gives 3.949 of average entropy per byte of encryption. That is 23% is improved than FABECS and 3% improved than CP-ABE-SD algorithm. Hence the proposed algorithm is out performed than the existed algorithms and frameworks in terms of PT of encryption and decryption and also average entropy per byte of encryption.

VI. Discussion

The Fig. 5 illustrates the performance comparison of three encryption algorithms, FABECS, CP-ABE-SD, and the proposed CBSA, in terms of encryption processing time (PT) across different file sizes ranging from 10MB to 90MB. The proposed CBSA demonstrates the lowest processing time across all file sizes. The processing time also increases linearly with the file size but at a much slower rate compared to FABECS and CP-ABE-SD. The proposed CBSA outperforms both FABECS and CP-ABE-SD, indicating a more efficient encryption process. The proposed CBSA is the most efficient encryption algorithm in terms of processing time, making it the preferred choice for handling large files.

The Fig. 6 illustrates the performance comparison of three decryption algorithms, FABECS, CP-ABE-SD, and the proposed CBSA in terms of decryption processing time (PT) across different file sizes ranging from 10MB to 90MB. The proposed CBSA demonstrates the lowest decryption processing time across all file sizes. The processing time increases linearly with the file size but at a slower rate compared to FABECS and CP-ABE-SD. This indicates that the proposed CBSA outperforms both FABECS and CP-ABE-SD, making it the most efficient decryption process.

The Fig. 7 compares the average entropy per byte for three different algorithms, FABECS, CP-ABE-SD, and the Proposed CBSA. The x-axis represents the different algorithms being compared, while the y-axis represents the entropy per byte, ranging from 0 to 5. Each bar represents the average entropy per byte for one of the algorithms. The entropy of FABECS per byte is slightly below 3. The entropy of CP-ABE-SD per byte is exactly 4. The entropy of the proposed CBSA per byte is also exactly 4.2. The FABECS algorithm has a lower entropy per byte compared to CP-ABE-SD and the Proposed CBSA. However, the proposed CBSA has better entropy per byte, which are higher than that of FABECS and CP-ABE-SD. The entropy per byte is a measure of the unpredictability or randomness of the data produced by each algorithm. Higher entropy generally indicates better security, as the data is more unpredictable. The Proposed CBSA has a higher entropy value, provides better security features compared to FABECS. Based on the average entropy per byte, the Proposed CBSA outperforms both CP-ABE-SD and FABECS in terms of the randomness and security of the data they produce.

VII. Conclusion

In this paper, an efficient framework named as confidentiality-based cloud storage is proposed and a confidentiality based security (CBS) algorithm is also developed to support that framework which optimizes the PT of both encryption and decryption procedures and assures integrity and confidentiality by using the degree of confidentiality based data classification. The CBS comprises classification algorithm CBC that classifies the data attributes based on confidentiality level, CBE algorithm used to encrypt and CBD algorithm used to decrypt the data based on their confidentiality level. Here, data is classified into three classes such as low-confidential, medium-confidential and high-confidential level data based on the degree of confidentiality (DC) value of different attributes calculated from the values given by the user, domain experts and other professionals. Then applied moderate security mechanism CP-ABE algorithm to medium-confidentiality level data and high security mechanism CP-ABE-SD is applied to high-confidential data, and RSA is applied to the low confidentiality data. So, processing cost for low confidential data is reduced more and processing cost for moderate-confidential data is also reduced. Hence, the data security process becomes optimal and cost effective in overall. The performance efficiency of our proposed algorithm has been

observed through the simulations conducted. The simulation results show that, our proposed framework achieved better PT for both encryption and decryption operations while assuring data integrity and confidentiality and the average entropy is also produced as better than existing algorithms FABECS and CP-ABE-SD.

In this paper the performance is evaluated in terms of only three parameters, as a future work some more parameters which improves the data security in the cloud.

REFERENCES

[1] AyadBarsoum and Anwar Hasan, "Enabling Dynamic Data and Indirect Mutual Trust for Cloud Computing Storage Systems", IEEE Transactions on Parellel and Distributed Systems, Dec. 2013 (vol. 24 no. 12), pp. 2375-2385.

[2] Pearson S, "Taking account of privacy when designing cloud computing services", Software Engineering Challenges of Cloud Computing, pages, 44 – 52, Vancouver, BC,2009.

[3] Kamara S, Lauter K, "Cryptographic cloud storage, Lecture Notes" in Computer Science 2010;6054:136–49.

[4] Pravin O. Balbudhe, Pradip O. Balbudhe; Cloud Storage Reference Model for Cloud Computing; International Journal of IT, Engineering and Applied Sciences Research (IJIEASR); Vol. 2, No. 3, pp.83, 2013.

[5] Rao, M. Varaprasad. "Data Duplication Using Amazon Web Services Cloud Storage." Data Deduplication Approaches,AcademicPress, 29 Jan. 2021.

[6] Frank Simorjay, "Data classification for cloud readiness" Microsoft Trustworthy Computing, 2014 Microsoft Corporation.

[7] Frank Simorjay, "Data classification for cloud readiness" Microsoft Trustworthy Computing, 2014 Microsoft Corporation.

[8] Wu J, Ping L, Ge X, Wang Y, Fu J; Cloud Storage as the Infrastructure of Cloud Computing, International Conference on Intelligent Computing and Cognitive Informatics (ICICCI),pp.383; 22-23 June 2010.

[9] Vengala, Dilip Venkata Kumar, D. Kavitha, and AP Siva Kumar. "Secure data transmission on a distributed cloud server with the help of HMCA and data encryption using optimized CP-ABE-ECC." Cluster Computing 23.3 (2020): 1683-1696.

[10] Sandeep K.Sood, "A combined Approach to Ensure Data Security in Cloud Computing" Journal of Network and Computer Applications 35 (2012) 1831–1838.

[11] M. A. Zardari, L. T. Jung and N. Zakaria, "K-NN classifier for data confidentiality in cloud computing," 2014 International Conference on Computer and Information Sciences (ICCOINS), 2014, pp. 1-6.

[12] Ming Li, Shucheng Yu, Yao Zheng, KuiRen and Wenjing Lou, "Scalable and Secure Sharing of Personal Health Records in Cloud Computing using Attribute-based Encryption", IEEE transaction on parallel and distributed systems, pages 131-43 vol. 24, issue 1, 2012.

[13] Yuan Cheng, Jaehong Park and Ravi Sandhu, Preserving User Privacy from Third-party Applications in Online Social Networks, Proceedings of the 22nd international conference on World Wide Web Companion, Pages 723-728. Geneva, Switzerland, 2013.

[14] Sergio Donizetti Zorzo, Rodrigo Pereira Botelho, Paulo Muniz de Ávila, Taxonomy for Privacy Policies of Social Networks Sites, Published Online, Social Networking, 2013, 2, 157-164 October 2013.

[15] N. Srinivasu, O. SreePriyanka, M. Prudhvi and G. Meghana," Multilevel classification of security threats in cloud computing", International Journal of Engineering & Technology, 7 (1.5) (2018) 253-257.

[16] Sudarsa, D. et al. "Enhanced data security through deep data classification in the cloud computing". International Journal of Emerging Trends in Engineering Research 8. 9(2020): 6226-6333.

[17] R. Kour, S. Koul and M. Kour, "A Classification Based Approach For Data Confidentiality in Cloud Environment," 2017 International Conference on Next Generation Computing and Information Systems (ICNGCIS), 2017, pp. 13-18.

[18] K. P. Singh, V. Rishiwal and P. Kumar, "Classification of Data to Enhance Data Security in Cloud Computing," 2018 3rd International Conference On Internet of Things: Smart Innovation and Usages (IoT-SIU), Bhimtal, India, 2018, pp. 1-5.

[19] O. Arki, A. Zitouni and A. Hadjali, "A Cloud Data Classification Model Using Fuzzy Logic," 2020 International Conference on Advanced Aspects of Software Engineering (ICAASE), 2020, pp. 1-6.

[20] Rizwana Shaikh, M. Sasikumar,"Data Classification for Achieving Security in Cloud Computing" , Procedia Computer Science,Volume 45,2015,Pages 493-498,ISSN 1877-0509.

[21] Ahamad, Danish, et al. "A Multi-Objective Privacy Preservation Model for Cloud Security Using Hybrid Jaya-Based Shark Smell Optimization." Journal of King Saud University - Computer and Information Sciences, 2020, https://doi.org/10.1016/j.jksuci.2020.10.015.

[22] Tawalbeh, Lo'ai, et al. "A Secure Cloud Computing Model Based on Data Classification." Procedia Computer Science, vol. 52, 2015, pp. 1153–1158.

[23] Vengala, Dilip Venkata Kumar, D. Kavitha, and AP Siva Kumar. "Three factor authentication system with modified ECC based secured data transfer: untrusted cloud environment." Complex & Intelligent Systems (2021): 1-14.

[24] Kartit, Z. et al. (2016). Applying Encryption Algorithm for Data Security in Cloud Storage. In: Sabir, E., Medromi, H., Sadik, M. (eds) Advances in Ubiquitous Networking. UNet 2015. Lecture Notes in Electrical Engineering, vol 366. Springer, Singapore. https://doi.org/10.1007/978-981-287-990-5_12.

[25] M. Morales-Sandoval, M. H. Cabello, H. M. Marin-Castro and J. L. G. Compean, "Attribute-Based Encryption Approach for Storage, Sharing and Retrieval of Encrypted Data in the Cloud," in IEEE Access, vol. 8, pp. 170101-170116, 2020.

[26] M. Thangavel and P. Varalakshmi, "Enabling Ternary Hash Tree Based Integrity Verification for Secure Cloud Data Storage," in IEEE Transactions on Knowledge and Data Engineering, vol. 32, no. 12, pp. 2351-2362, 1 Dec. 2020.

[27] Fursan Thabit, Sharaf Alhomdy, Sudhir Jagtap, A new data security algorithm for the cloud computing based on genetics techniques and logical-mathematical functions, International Journal of Intelligent Networks, Volume 2, 2021, Pages 18-33, ISSN 2666-6030.

[28] O. Arki, A. Zitouni and A. Hadjali, "A Cloud Data Classification Model Using Fuzzy Logic," 2020 International Conference on Advanced Aspects of Software Engineering (ICAASE), Constantine, Algeria, 2020, pp. 1-6.

[29] A. Yeboah-Ofori, S. K. Sadat and I. Darvishi, "Blockchain Security Encryption to Preserve Data Privacy and Integrity in Cloud Environment," 2023 10th International Conference on Future Internet of Things and Cloud (FiCloud), Marrakesh, Morocco, 2023, pp. 344-351.

[30] P. Swathika and J. R. Sekar, "Role-based Access and Advanced Encryption Techniques Ensure Cloud Data Security in Data Deduplication Schemes," 2023 7th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Kirtipur, Nepal, 2023, pp. 225-232.

[31] R. R. Prasad and A. Kumari, "Cloud Data Security using Balanced Genetic Algorithm," 2023 9th International Conference on Electrical Energy Systems (ICEES), Chennai, India, 2023, pp. 132-137.

[32] M. M. R and A. T.P, "Novel Weight-Improved Particle Swarm Optimization to Enhance Data Security in Cloud," 2023 7th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Kirtipur, Nepal, 2023, pp. 195-200.

[33] S. R and M. S. L. Devi, "Fragment Security Framework for Enhancing Data Security in Cloud Services," 2023 Third International Conference on Digital Data Processing (DDP), Luton, United Kingdom, 2023, pp. 205-210.

[34] N. Dwivedi, M. Swarnkar, A. Soni and M. Singh, "Cloud Security Enhancement Using Modified Enhanced Homomorphic Cryptosystem," 2023 IEEE Renewable Energy and Sustainable E-Mobility Conference (RESEM), Bhopal, India, 2023, pp. 1-6.

[35] P. Yellamma, C. Narasimham and V. Sreenivas, "Data security in cloud using RSA," 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), 2013, pp. 1-6.

[36] J. Bethencourt, A. Sahai, and B. Waters, “Ciphertext-policy attribute-based encryption,” in Proc. IEEE Symp. Secur. Privacy, 2007, vol. 2008, pp. 321–334.

[37] N. Chen, J. Li, Y. Zhang and Y. Guo, "Efficient CP-ABE Scheme With Shared Decryption in Cloud Storage," in IEEE Transactions on Computers, vol. 71, no. 1, pp. 175-184, 1 Jan. 2022.

[38] https://www.techtarget.com/searchsecurity/definition/Transport-Layer-Security-TLS.

[39] S. Tsang, B. Kao, K.Y. Yip, W.-S. Ho, and S.D. Lee, “Decision Trees for Uncertain Data,” Proc. Int’l Conf. Data Eng. (ICDE), pp. 441-444, Mar./Apr. 2009.

[40] D. Boneh and M.K. Franklin, "Identity-Based Encryption from the Weil Pairing", Proc. Intl Cryptology Conf. Advances in Cryptology, 2001.

[41] A. Sahai and B. Waters, "Fuzzy identity-based encryption" in Advances in Cryptology, Berlin, Germany:Springer-Verlag, vol. 3494, pp. 457-473, 2005.

[42] V. Goyal, O. Pandey and A. Sahai, "Attribute-based encryption for fine-grained access control of encrypted data", Proc. 13th ACM Conf. Comput. Commun. Secur. (CCS), pp. 89-98, Oct./Nov. 2006.

[43] Rao, J. & Reddy, Vuyyuru & Hota, Prakash. (2020). Enhanced Ciphertext-Policy Attribute-Based Encryption (ECP-ABE). 10.1007/978-981-13-8461-5_57.